PATENT ABSTRACTS OF JAPAN

(11)Publication number:

2002-108844

(43) Date of publication of application: 12.04.2002

(51)Int.CI.

G06F 17/21 G06F 12/00

(21)Application number: 2000-295352

(71)Applicant: HITACHI SOFTWARE ENG CO LTD

(22)Date of filing:

28.09.2000

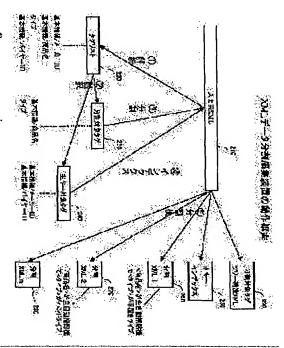
(72)Inventor: KIJIMA NORIKAZU

(54) XML DATA DIVISION EDITING DEVICE

(57) Abstract:

PROBLEM TO BE SOLVED: To quickly retrieve and efficiently edit a large quantity of XML data by dividing them.

SOLUTION: Input source XML data 110 are analyzed to prepare a tag list 120. A main key index 150 for corresponding a main key to the split XML data, split XMLs 160, 170 and 180 divided by a split object tag and a split object tag tree structure XML 190 are prepared from the tag list 120 by using the split object tag 130, a main key object tag 140 and the data 110.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

(18) 日本国特部庁 (JD)

(ID) 公開特許公報(A)

(11)特許出願公開番号

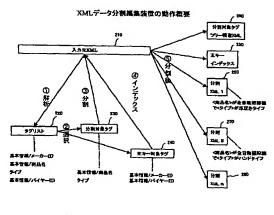
(P2002-108844A) (43)公開日 平成14年4月12日(2002.4.12) 4 8 4 特開2002-108

(51) Int. C1.7		数別配母	F 1		テー7コード (参考)
G 0 6 F	17/21	501	GOGF	17/21 501 2	58009
		570		570 2	58082
	12/00	547		12/00 547 H	~
	路效點次	13次 未請求 請求項の数4	0 L	쇗	(全13頁)
(21) 出版番号	48	特额(2000-295352 (P2000-295352)	(71) 出版人 000233055	000233055	
				日立ソフトウエアエ	日立ソフトウエアエンジニアリング株式会
(22) 出版日	邓成	平成12年9月28日 (2000. 9. 28)		社	
				种奈川県横浜市中区	种奈川県横浜市中区區上町6丁目81番地
			(72) 発明者	木島 教和	
				神奈川県横浜市中区	种奈川県橫低市中区尾上町6丁目81番地
				日立ソフトウエアエ	日立ソフトウエアエンジニアリング株式会
				社内	
			(74) 代理人	100073760	
				并型士 鈴木 誠	
			トターム(数	Fターム(容表) 5B009 QA!! SA03 VA02	13 VA02
				5B082 EA01 GA08	82

(54) 【発明の名称】XMLデータ分割編集装置

(57) [更約]

(解決手段) 入力元XMLデータ110を解析してタ グリスト120を作成し、タグリスト120から、分割 対象タグ130と主キー対象タグ140と入力元XML 110を使用して、主キーと分割XMLデータを対応付 ける主キーインデックス150、分割対象タグで分割さ れた分割XML160,170,180、分割対象タグ 【磔蹈】 大田なXMLデータを分割することによっ て、迅速に検索し、効率的に確集できるようにする。 ツリー街道XML 190を作成する。



[特許館水の範囲]

【請求項1】 XMLデータを編集・検索する装置であ タグと分割対象タグを用いて前記XMLデータを分割す って、入力元XMLデータのタグとタグの値に基づい

躓において、入力元XMLデータについて、1レコード 請求項1記載のXMLデータ分割編集装 中の同階層に1つだけ含まれるタグを対象にして、同じ タグ値を収る同名タグとタグの出現頻度を計測し、タグ リストを生成することを特徴とするXMLデータ分割編

【請求項3】 請求項1,2 記載のXMLデータ分割編 を特徴とするXMLデータ分割幅集装置。 【指水項4】 調水項3記銭のXMLデータ分割編集数 **当において、 ±キーインデックスと分割対象タグツリー** 構造を分割XMLデータの検索に使用することを特徴と するXMLデータ分割編集装置。

置に係り、特に大規模なXMLデータの編集、検索など 【発明の属する技術分野】本発明はXMLデータ編集装 に好適なXMLデータ分割編集装置に関する。

[0002]

【従来の技術】XMLデータを編集する場合、一つのX

え、処理装置10の処理結果としての主キーインデック 80、及び分割対象タグツリー構造XML 190などを 格納する。なお、入力元XMLデータ110の配位媒体

ス150、複数の分割XMLデータ160, 170,

승 森、塩塩する方法については、スタイルシートやクエリ ーといった手法がよく知られているが、一つのXMLデ ータのみ対象としており、一つのXML内に複数のXM Lを読み込むようなXMLであってもすべてのXMLを 順に読み込む必要があり、XMLデータを効率的に編集 **下ることは困癖である。**

(0004)

20 をデータとして使用する場合、データ量が少ない場合は 問題ないが、データ型が肥大化した場合、検索、更新等

特開2002-108844

ව

る。肥大化したXMLデータを効率的に扱うためにはX MLデータを分割する必要がある。

[0005] 本発明の目的は、上記従来技術の問題点を 将沙し、大規模なXMLデータの効率的な編集、检索等 を実現するXMLデータ分割編集装置を提供することに

[9000]

【歌蹈を解決するための手段】上記目的を遠成するため に、本発明は、入力元XMLデータを、そのタグとタグ の値の出現頻度などに基づいて解析してタグリストを生 成する手段と、核タグリストから選択した主キー対象タ グと分割対象タグを用いて入力元XML データを分割す る手段を備えることを主要な特徴とする。 2

グを対象にして、同じタグ値を取る同名タグとタグの出 **悩を持ち、1レコード中の同階圏に1つだけ含まれる夕** 現気度を計測し、タグリストを生成する。このタグリス トから選択したタグに基づいて、ツリー結婚のディレク トリを作成し、複数のXMLファイルへ分割する。XM レデータ上の主キーとなる項目については更新をすばや く行うために虫キーと格徴XMLファイルを対応づける [0007] 本発明では、XMLデータ中に含まれる。 インデックス用XMLファイルを作成する。 20

[0008]

(発明の実施の形態)以下、図面により本発明の一実施 の形態について説明する。図1は、本発明によるXML 0、数示数置20、キーポード30、マウス40、モデ クなどの外配値装置70などから構成されるが、このハ 本XMLデータ分割編集装置は、処理装置(CPU)) ム50、RAMなどの記憶装留60、及びハードディス **ードウェア結成自体は所聞パンコンやワークステーツョ** ンなどと基本的に同様である。ここで、処理数徴10は 本発明に関係する手段としてタグリスト生成手段11と XMLデータ分割手段12を具備している。配位装置6 0はタグリスト120、分割対象タグ130、主キー対 は、あらかじめ取得した入力元XMLデータ110に加 データ分割編集装置の一実施形態のプロック図である。 象タグ140を一時的に格納する。外部記憶装置70 30

【0009】図2は本×MLデータ分割幅集装置の動作 既要を示した図である。処理装置10のタグリスト生成 手段11は、外部記憶装置70から入力元XMLデータ 110を取り込み、核入力元XMLデータ110をタグ の出現頻度で解析してタグリスト120を生成し、記憶 数型120に格納するとともに扱示装置20に表示す と処理結果を格納する記憶媒体とは別構成でもよい。

る。ユーザは、キーボード30やマウス40により、投 示装置20に表示されたタグリスト120から分割対象

て、該入力元XMLデータを解析してタグリストを生成 する手段と、前記タグリストから選択された主キー対象 る手段とを有することを特徴とするXMLデータ分割幅

[請水項2]

集装置において、XMLデータを分割する際に、主キー して略層化した分割対象タグツリー構造を作成すること 対象タグの値と分割XMLデータとの関連付けを示した **士キーインデックスと、分割対象タグの値をタグの値と**

[発明の詳細な説明] [0001]

ることができない。また、大街のXMLデータの中から 目的のデータを見つけることが困難である。このような 理由から、XMLデータ監が少ない場合はXMLデータ を協筑、依案することが難しくないが、XMLデータ費 MLファイルのデータをすべて読み込まなければ猫集す が増えると、XMLデータを編集することが非常に困難 [0003] 従来、XMLデータをすべて読み込み、敬

に非常に時間がかかる上、目で確認することが困難であ (発明が解決しようとする課題] 従来技術では、XML

40は記憶装置60に格納される。処理装置10のXM 0の値をタグの値として階層化した分割対象タグツリー タグ130と主キー対象タグ140を遊択する。この分 を、分割対象タグ130によって分割して分割XMLデ 対象タグ140の値と該当レコードが収められた分割X MLデータ160、170、180との関連付けを示し た主キーインデックスXML150、分割対象タグ13 **構造XML190を生成し、外部配位装置70に格約す** る。なお、タグリスト120なども外部記憶装置70に 割対象タグツリー構造XML130と主キー対象タグ1 **一タ160、170、180を生成し、同時に、主キー** Lデータ分割手段12は、入力元XMLデータ110 以持しておけば、あとで再利用が可能である。

【0010】以後、ユーザから検案要求や幅集要求等が あると、処理数置10では、主キーインデックス150 をもとに分割対象タグツリー構造XML 190を表示数 置20に表示し、ユーザが分割対象タグを選択すると、 該当分割XMLデータを外部記憶装置70から競み出し て表示装置20に表示する。 [0011] 図7に入力元XML110の具体例 (入力 710には、このような商品データが多数収められてい ル710は母初の前品データ720を含んでいる。 滴品 4、「消費電力」タグ725、「価格」タグ726を持 示す。図10は分徴対象タグツリー格温XML190の 元XMLファイル710)を示す。入力元XMLファイ **報」タグ721は子ノードに「メーカー1D」タグ72** 2と「バイヤー1D」タグ723、「商品名」タグ72 7、「単位」タグ728を持つ。入力元XMLファイル (主キーインデックスXML810)を示す。 図9は分 斑XML (1) 160の具体例 (分割XML910) を 具体例(分割対象タグツリー構造XML 1010)を示 データ720は「商品」タグの子ノードに「基本情報」 タグ721と「タイプ」タグ729を持つ。「基本情 つ。「価格」タグ726は子ノードに「値」タグ72 る。 図8 は主キーインデックス XML 150 の具体例

[0013] 図3はタグリスト生成手段11の処理フロ がら本XMLデータ分割編集装置のタグリスト生成手段 **ーチャートである。このフローチャートに従って、煎品** 【0012】以下、図7乃至図10の具体例を参照しな データが収められた XMLファイル7 1 0 からタグリス 11とXMLデータ分割手段12の動作を詳述する。 トを生成する処理を以下に示す。

ある。 商品タグで囲まれた部分が一つの商品データを投 [0014] タグリスト生成手段11は、まず、一つの ートタグ「document」の直下に「商品」タグが 入力元XMLファイル710に着目する(ステップ30 1)。 そして、1レコードを没すタグがまず存在するか 判定する (ステップ302)。 図7に示す入力元XML ファイル710には液晶のデータが収められており、ル

したおり、17コードとみなす。図ったは、一つの短距 (レコード)を設すタグがまだ存在する。そこで、一つ の商品データ(1レコード)720に着目し(ステップ こで、一つのタグ「基本情報」721を取り出す(ステ 303)、その中にまだ取り出していないタグがまだ符 在するか判定する (ステップ304)。 1商品データ7 20の中にはまだ取り出していないタグが存在する。そ ップ305)。そして、この取り出したタグは値を持

ち、同階層タグに同名タグが存在しないか判定する(ス テップ306)。「基本情報」タグ721は値を持たな (ステップ305)。 [基本債報/メーカー1D1タグ い。そこで、ステップ304に戻り、再び1商品データ 次のタグ「基本情報/メーカーID」722を取り出す い、そこで、タグの名前「基本情報/メーカー1D」? メーカー101722のタグの数を一つインクリメント し、「基本債糧/メーカー1D」タグ722の名前の数 は1となる (ステップ307)。また、「基本俯靼/メ ーカーID」タグ722の値「M0001」の数を一つ インクリメントし、「基本値報/メーカーID」タグ7 2 2 の値「M 0 0 0 1」の数は1 となる (ステップ 3 0 の中にまだ取り出していないタグが存在するか判定し、 22とタグの値「M0001」を取得し、「基本情報/ 7.2.2は値を持ち、タグの同路層に同名タグを持たな

【0015】次に、ステップ304に戻り、「基本情報 ノバイヤー1D」タグ723、「基本情報/商品名」タ ガ724、「基本情報/消毀電力」タガ725、「基本 位」タグ728、「タイプ」タグ729の順に処理を繰 信報/信格/値」タグ727、「基本位報/恒格/単 り返し、一つの商品データ720の解析を終了する。

て、入力元XMLファイル710にある商品の終わりま で繰り返す。この結果、入力元XMLファイル710の **会感品 (金レコード) について、1レコード中のタグ中** でタグの同階層に一つのみ出現するタグの出現頻度と同 [0016]次に、ステップ302に戻り、同様にし 名タグで同じ歯の数が全て解析される。

る。「タイプ」タグの数は100個、「タイプ」タグの 値が「床盥きタイプ」の数は60個、「タイプ」タグの 商品名」タグの数は100個、「基本情報/商品名」タ [0017] ここで、図7の入力元XMLファイル71 値が「ハンドタイプ」の数は40個ある。「基本情報ン グの値が「全自動期除機」の数は40個、「基本情報/ 0の金商品データの解析結果は次のようになったとす 商品名」タグの値が「全自動掃除機」の数は35個、

\$

消費電力」タグの値が「200W」のタグの数は15個 「基本情報/商品名」タグの値「全自動冷磁庫」の数は 25個ある。「基本情報/消費電力」タグの値が「50 W」のタグの数は65個、「基本情報/消費電力」タグ の値が「300W」のタグの数は20個、「基本情報/

20

[0018] 図4はタグリストの具体例であり、図7の 入力元XMLファイル710の全商品データの解析結果 を上記のように仮定して、同名タグの数の多い類、同名 グリスト410は、入力XMLデータ710中に含まれ る1レコード (1 商品データ) 中に一つだけ出現するタ グ名412が入力XMLデータ710中に含まれる同名 タグ数413、同タグ名で同じタグの値の数414の大 きい顔に順位411付けして並べたリストである。 ここ では、同タグ名で同じタグの値の数414は同名タグが 同じ値をとる数を大きい順に数番目まで示し、タグリス ものである。図4中、410がタグリストを示す。該タ ト410では4番目まで出力されている。タグの値41 タグで同じ値の数の多い順にタグ名とタグの値を並べた 5 は同名タグがとるタグの値を数個出力する。

が60個、「ハンドタイプ」が40個あるので、同タグ 帝目に多いのが40個となり、タグの値415には「床 の数は100個なので岡名タグ数413は100個、入 [0019] 例えば、入力XMLファイル710中に含 まれ、1レコード中に一つだけ存在する"タイプ"タグ 名で同じタグの値の数414は1番多いのが60個、2 カXMLファイル710中に含まれ、1レコード中に一 つだけ存在する「タイプ」タグの値は「味置きタイプ」 **踏きタイプ、ハンドタイプ」と出力される。**

8 【0020】ユーザは、タグリスト410から分割に適 切と思える分割対象タグ名420を選択する。また、タ グリスト410から主キーとして使用する主キー対象タ グ名430を遊択する。ここでは、分割対象タグ名42 タグを遊択する。主キー対象タグ名430としては「基 本情報/メーカーID」と「基本情報/バイヤーID」 0として「基本情報/商品名」「タグと」と「タイプ」 タグを選択する。

ローチャートである。以下、図5のフローチャートに従 い、入力元XMLとして図7の入力元XML710を使 [0021] 図5はXMLデータ分割年段12の処理フ 名」タグと「タイプ」タグを指定し、主キー対象タグと して同じく図4の「基本情報/メーカー1D」と「基本 情報/パイヤーID」タグを指定した場合について説明 用し、分割対象タグ名として図4の「基本情報/商品

[0022] XMLデータ分割手段12は、まず、入力 4)。1前品720を表すタグ群の中でまだ籍目してい て、まだ着目していないレコードは存在するか判定する タグを1レコードとして取り扱う。入力元XML710 こで、入力元XML710において1レコード目の商品 3) ・そして、1レコード (1 商品データ) 中にまだ着 **目していないタグは存在するか判定する(ステップ50** (ステップ502), 入力元XML710では「商品」 データ720を表すタグ群に着目する (ステップ50 元XML710に着目する (ステップ501) 。そし にまだ落目していない商品データ720が存在する。

特限2002-108844 [基本情報] タグ721に着目す

3

のタグが判定する (ステップ506)。 「基本付報」タ **グ721は主キーのタグではない。そこで、次に着目す** る (ステップ505)。 そして、俗目するタグは主キー るタグは分割対象タグか判定する (ステップ508)。 「基本情報」タグ721は分割対象タグではない。 ないタグが存在する。

[0023] ステップ504に戻り、1レコード (1前 品データ)中にまだ着目していないタグは存在するか判 定する。商品720を表すタグ群の中で発目していない - I D」タグ722は主キーを殺すタグである(ステッ プ506)。「基本情報/メーカー1D」タグ722の 07)。 [基本情報/メーカーID]タグ722は分割 タグは存在する。「基本情報/メーカーID」タグ72 2に第目する (ステップ505)。 [基本憐穀/メーカ 名前とタグの値「M0001」を保持する (ステップ5 お象タグではない(ステップ508)。

[0024] 再びステップ504に戻る。前品710を **数すタグ群の中で着目していないタグは存在する、「基** 本情報/バイヤー1D」タグ723に着目する (ステッ プ505)。 [基本債報/パイヤー! D」タグ723は 主キーを没すタグである (ステップ506)。「基本情 報/バイヤーID」タグ723の名信とタグの道「B0 0 0 1 」を保持する (ステップ5 0 7) 。 「基本情報人 バイヤー 1 D」タグ723は分割対象タグではない(ス テップ508)

ップ506)。「基本情報/底品名」タグ724は分割 商品名」タグ724は主キーを表すタグではない(ステ 対象タグである (ステップ508), そこで、「基本情 報/商品名」タグ724の名前と値「全自動掃除機」の [0025] 再びステップ504に戻る。商品720を **支すタグ群の中で着目していないタグは存在する。「基** 本情報/商品名」タグ724に着目する。「落本情報/ 値を保持する (ステップ509)。

[0026] 同様にして、1商品データ720を改すタ ゲ群の残りのタグについても、ステップ504~509 グ、分割対象タグとして現れるものは、分割対象タグと してー「タイプ」タグ129があり、「タイプ」タグ1 に基づいて処理する。残りのタグの中で主キー対象タ 29の値は「床置きタイプ」である。

[0027] XMLデータ分割手段12は、1所品を投 すタグ群の中で着目していないタグが存在しなくなった 場合、 主キー対象タグ名及びタグの磁、並びに分割対 象タグ名及びタグの値を主キーインデックスXML15 0に沓き込む (ステップ510)。 主キーインデックス XML150は主キー対象タグと格納先XMLとの関連 49

[基本情報/メーカー1D」タグ122とタグの値 [M 00011、「基本情報/バイヤー1口」タグ123と タグの超「B0001」である。分割対象タグ名、タグ [0028] ここでは、主キーのタグ名、タグの倒は

20

特開2002-108844

0を使用して上記例のように分割した場合、 出キーイン 10のように蛰き込まれる。図8において、主キーイン デックスXML810は「document」タグの子 タグは「主キー対象」タグ821、「分割対象」タグ8 デックスXML150は、土キーインデックスXML8 26を子ノードとして持つ。「主キー対象」タグ821 [0029] 図8に示したように、入力元ファイル71 [主キー対象] タグ821は「基本情報」タグ821を 持つ。「基本情報」タグ821は「メーカー1D」タグ 824と「バイヤーID」タグ825を持つ。「分割対 4g」タグ826は分割XMLの名前を指定するために必 **一ドとして「基本情報」タグ827と「タイプ」タグ8** 29を持つ。「基本情報」タグ821は「商品名」タグ ノードとして「商品」タグ820を複数持つ。「商品」 要な情報が記録される。「分割対象」タグ826は子ノ は入力元XMLの主キー対象タグの情報が記録される。 828を持つ。

2

[0030] 次に、XMLデータ分割手段12は、分割 タグの値を名前にした XMLは作成されているか判定す Lは作成されていない。そこで、分割対象タグの値「全 512)。そして、分割タグの値が分割タグツリー構造 では、分割タグの値が分割タグツリー構造XML190 に存在しない。そこで、分割タグツリー構造XML19 「会自動部除機」と「埰選者タイプ」を名前にしたXM [全自動植除機羊床置きタイプ] を作成する (ステップ XMLに存在するか判定する (ステップ513)。 ここ 倍回している1 協品データ810を没すタグ群を分割X ML「全自動桶除機羊床置きタイプ」に書き込む(ステ 0 にタグの値を追加する (ステップ514)。そして、 る (ステップ511)。 ここでは、分割対象タグの値 自動掃除機」と「床躍きタイプ」を名前にしたXML

[0031] その後、ステップ502に戻り、XMLデ ータ分割手段12は、同様の処理を入力元XML710 の本へたの混品ゲータにしてた数の版を。 7515).

に分割XML 「全自動樹除機羊尿脳きタイプ」910を 示している。また、図10は分割対象タグツリー構造X [0032] 図9は分割XML160の具体例であり、 ML190の具体例を示している。

[0033] 図6に、分割後XMLを使用して編集、検 深等を行う場合の具体例を示す。 本XMLデータ分割編 する。週択した分割対象タグ名624に基づいて、分割 [基本情報/バイヤー I D J の値622から歯品を特定 され分割対象タグの値を使用した分割対象タグツリー格 集装器は、主キーインデックスXML150を使用し て、主キー「基本情報/メーカー1D」の値621と

集部分に該当する分割XML626が表示され、幅集機 造615を表示する。ユーザが分割対象タグツリー構造 625で指定した分割対象タグを選択すると、XML編 能が提供される。さらにタグの値一覧指定タグ627に てタグを指定し、タグの値一覧表示628を選択すると タグの値一覧629にタグの値一覧が表示される。タグ 編集部分に該当するXML626内の該当する商品の部 の値一覧選択629にてタグの値を選択すると、XML 分が表示される。

となったタグについては高速に枚索する事ができ、目で 確認することが容易になる。また、XMLファイルを解 り、協き込みすることができるようになる。タグの出現 分割されたXMLファイルを検案する場合は、分割対象 [発明の効果] 以上説明したように、本発明によれば、 **析する場合、すべしの内容を読み込む必要があるので、** 類度に基づいた検索キーによるXMLの分割によって、 XMLファイルのサイズが小さけたば、南海に競争政 遊やかに餃祭、編集することができるようになる。 [0034]

【図1】 本発明によるXMLデータ分割編集装置の構成 例を示すプロック図である。 【図面の簡単な説明】

ន

[図2] 本発明によるXMLデータ分割編集装置の動作 既要を説明する図である。

[図3] タグリスト生成の処型フローチャートである。

[図4] タグリストの異体例を示す図である。

[図5] XML分割の処理フローチャートである。

[図6] 分割XMLデータ利用の具体例を示す図であ

【図8】 主キーインデックス X M L の具体例を示す図で [図7] 入力元XMLデータの具体例を示す図である。

[図10] 分割対象タグツリー構造XMLの具体例であ 【図9】分割XMLデータの異体例を示す図である。

[符号の説明]

10 処理数階

XMLデータ分割手段 タグリスト生成手段

60 記憶装置 20 表示裝置 49

110 入力元XMLデータ 70 外部記憶装置

120 97UXF

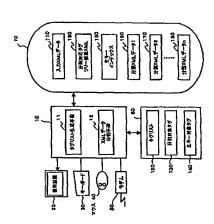
40 主キー対象タグ 分割対象タグ 130

主キーインデックスXML 150

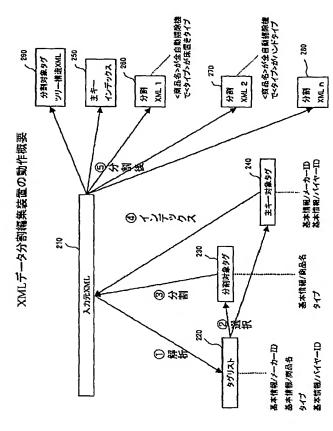
160、170、180 分割XMLデータ 190 分割対象タグツリー構造XML

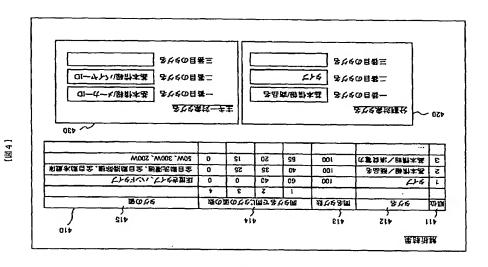
(図]

XMLデータン型指集技能

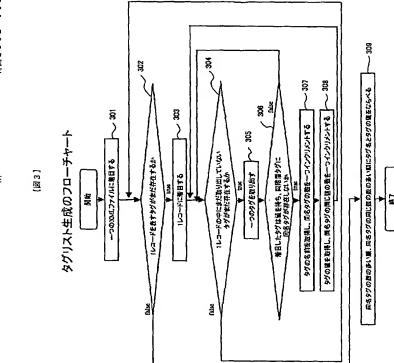


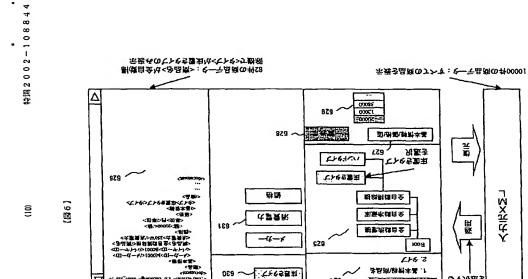
[図2]





421146





1000M : CI--C--

+ 年 年

キ主の中来訴

例本具式J用東多置装集融階代や一下JMX

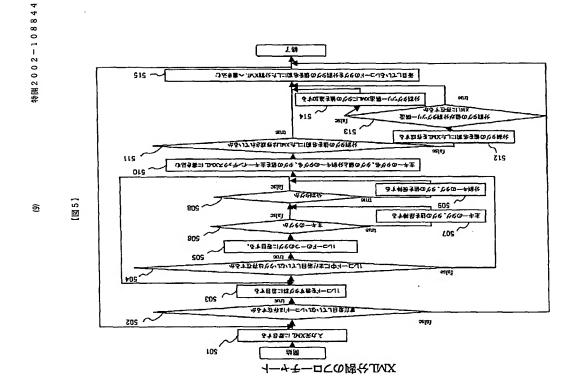
029

またた→田代小り別店

-+∓

イスいやを るで用班字

ST WX TYY



(13)

Ξ

[四]

入力元XMLデータの具体例

[88]

主キーインデックスXMLの具体例

[図10]

分割対象タグツリー構造 XMLの具体例

) 1010

 <(7xml version="1.0" encoding="Shift_INS"?>

 <全自動光深機>

 <全自動光深機>

 <全自動光液体>

 <全自動光液体>

 <全自動光液体>

 <全自動光液体>

 <<全自動光液体>

 <</td>
 <</td>

 <</td>
 <</td>

 <</td>
 <</td>

 <</td>

[6図]

分割後XIML 1の具体例

